Paper 1

Abstract Title Page

Title:

A General Framework for Power Analysis to Detect the Moderator Effects in Two- and Three-Level Cluster Randomized Trials

Authors and Affiliations:

Nianbo Dong
University of Missouri

Jessaca Spybrook
Western Michigan University

Ben Kelcey
University of Cincinnati

Abstract Body

Background / Context:

Like studies that focus on detecting a main effect, a critical consideration in designing studies to detect moderation effects is the statistical power with which the moderation effect can be detected if they exist. The potential moderators of the intervention in cluster randomized trials (CRTs) include pretest, ethnicity, school climate, or the fidelity of implementation, which could be at different levels and have different distributions (e.g., binary, continuous). For moderator relationships in experimental studies, Bloom (2005), Spybrook (2014), and Spybrook, Kelcey, and Dong (2015) have presented procedures for conducting power analysis for binary moderators in two- to four-level cluster randomized trials (CRTs), but have not extended those procedures to include continuous moderator variables. Furthermore, no computational tools to facilitate use of these techniques by researchers have been developed.

Most recently, Mathieu, Aguinis, Culpepper, & Chen (2012) conducted a comprehensive Monte Carlo simulation to estimate the statistical power to detect cross-level interaction effects, and Dong (2014) presented the formulas to calculate minimum detectable effect size (MDES) for continuous moderator analysis in two-level CRTs with a level-2 continuous moderator. However, Mathieu et al (2012) only studied two-level analyses without including covariates, and did not provide closed form formulas to estimate the statistical power, MDES, or minimum required sample size to detect meaningful effects; Dong (2014) only focused on a level-2 continuous moderator in two-level CRTs, and didn't study cross-level moderation, include covariates, or examine three-level CRTs. In sum, there is no a systematic study investigating the power analysis to detect moderation effects in two- and three-level CRTs that includes both binary and continuous moderators, same and cross-level moderation, with covariates, and computational tools.

Purpose / Objective / Research Question / Focus of Study:

The purpose of this study is to propose a general framework for power analyses to detect the moderator effects in two- and three-level CRTs. Specifically, we aim to: (1) develop the statistical formulations for calculating statistical power, minimum detectable effect size (MDES) and its confidence interval to detect the moderation effects in two- and three-level CRTs, which include same and cross-level moderation, binary and continuous moderators, and covariates, and (2) operatize these formulas in the enhanced version of *PowerUp!* (Dong & Maynard, 2013) to create spreadsheets for calculating power, MDES, etc.

Significance / Novelty of study:

Educational researchers have interests in the effects of both binary and continuous moderators in CRTs. Statistical power analysis is appropriate in the planning stages to help researchers design studies with sufficient power to detect such relationships when they are large enough to have practical or theoretical significance. This study will provide a general framework and computational tool for power analyses to detect the binary and continuous moderator effect in two- and three-level CRTs.

Study Design:

This study covers the same and cross-level moderation in two- and three-level CRTs. The detailed study designs and analysis models are presented at Table 1 (see appendix). For the cross-level moderation, i.e., the moderator at the level lower than the treatment level, there are two options: the fixed slope and random slope of the moderator variable. The fixed slope of the moderator variable assumes that the effect of the moderator varies by the treatment status, but does not vary across the higher-level clusters, while the random slope of the moderator effects assume that the effect of the moderator varies by the treatment status, and varies randomly across the higher-level clusters. The power and MDES formulas are derived for all the designs and models covered in Table 1 and are applied in *PowerUp!*.

Statistical, Measurement, or Econometric Model:

Due to the page limitation, we present the results of two-level CRTs with a treatment variable at Level 2 and a moderator at Level 1 below.

The HLM, including one treatment variable, T_j, and one level-1 moderator, X_{ij}, with a *random* slope is:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$, $r_{ij} \sim N(0, \sigma^2_{|X})$ (1)

Level 2: $\begin{matrix} \beta_{0j} = \gamma_{00} + \gamma_{01} T_j + u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11} T_j + u_{1j} \end{matrix}$, $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2_{00|T} & \tau_{01|T} \\ \tau_{10|T} & \tau^2_{11|T} \end{pmatrix} \right)$ (2)

The interest for a moderator analysis is whether the parameter, γ_{11}, which indicates the conditional relationship between the average treatment effect and the moderator, is statistically significant.

According to Expression [3.89] in Raudenbush & Bryk (2002, p.59),

$$SE(\hat{\gamma}_{11}) = \sqrt{\frac{\tau^2_{11|T} + \sigma^2_{|X} / (n\sigma^2_X)}{P(1-P)J}} = \sqrt{\frac{(1-R^2_{2T})\omega\tau^2_{00} + (1-R^2_1)\sigma^2 / (n\sigma^2_X)}{P(1-P)J}} \quad (3)$$

where R^2_1 is the proportion of variance at level 1 that is explained by the level-1 moderator (X_{ij}): $R^2_1 = 1 - \sigma^2_{|X} / \sigma^2$. R^2_{2T} is the proportion of variance between level-2 clusters on the effect of X_{ij} explained by level-2 predictor (T_j): $R^2_{2T} = 1 - \tau^2_{11|T} / \tau^2_{11}$. $\omega = \tau^2_{11} / \tau^2_{00}$ indicates the effect heterogeneity for the level-1 covariate (X_{ij}) across level-2 units (clusters) in the model that is not conditional on treatment variable, T_j, which is the proportion of the variance between clusters on the effect of X_{ij} to the between-cluster residual variance. σ^2_X is the variance of X_{ij}, and P is the proportion of clusters in the treatment group.

The noncentrality parameter (unstandardized) is:

$$\lambda_{|X} = \sqrt{\frac{\hat{\gamma}^2_{11} P(1-P)J}{(1-R^2_{2T})\omega\tau^2_{00} + (1-R^2_1)\sigma^2 / (n\sigma^2_X)}} \quad (4)$$

By standardization, let $\tau^2_{00} + \sigma^2 = 1$ and $\sigma^2_X = 1$, the standardized coefficient $\delta = \gamma_{11}$, or let $\delta = \gamma_{011}\sqrt{\frac{\sigma^2_X}{\tau^2_{00} + \sigma^2}}$, the noncentrality parameter (standardized) is:

$$\lambda_{|X} = \sqrt{\frac{\delta^2 P(1-P)J}{(1-R^2_{2T})\rho\omega + (1-R^2_1)(1-\rho)/n}} \quad (5)$$

The degrees of freedom is $v = J - 2$. ρ is the unconditional intraclass correlation, $\rho = \tau^2_{00} / (\tau^2_{00} + \sigma^2)$.

The statistical power is: $1 - \beta = 1 - P[T^{'}(J-2, \lambda_{|X}) < t_0] + P[T^{'}(J-2, \lambda_{|X}) \le -t_0]$.

The minimum detectable effect size (MDES) regarding the standardized coefficient is:

$$MDES(|\delta|) = M_v \sqrt{\frac{(1-R_{2T}^2)\rho\omega + (1-R_1^2)(1-\rho)/n}{P(1-P)J}} \tag{6}$$

where, $M_v = t_\alpha + t_{1-\beta}$ for one-tailed tests with v degrees of freedom ($v = J-2$), and $M_v = t_{\alpha/2} + t_{1-\beta}$ for two-tailed tests. The 100*(1-α)% confidence interval for $MDES(|\delta|)$ is given by: $(M_v \pm t_{\alpha/2})\sqrt{\frac{(1-R_{2T}^2)\rho\omega + (1-R_1^2)(1-\rho)/n}{P(1-P)J}}$ (7)

Extension to the Fixed Slope Model

This *fixed* slope HLM assumes that the effect of X_{ij} varies by the treatment status (T_j), but does not vary across level-2 units as in the random slope model in Expression 2.

The Level 2 model is: $\begin{array}{l}\beta_{0j} = \gamma_{00} + \gamma_{01}T_j + u_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11}T_j\end{array}$, $u_{0j} \sim N(0, \tau_{|T}^2)$ (8)

The noncentrality parameter (unstandardized) is: $\lambda_{|X} = \sqrt{\frac{\hat{\gamma}_{11}^2 P(1-P)Jn\sigma_X^2}{(1-R_1^2)\sigma^2}}$ (9)

By standardization, let $\tau^2 + \sigma^2 = 1$ and $\sigma_X^2 = 1$, or let $\delta = \gamma_{11}\sqrt{\frac{\sigma_X^2}{\tau^2+\sigma^2}}$. The noncentrality parameter (standardized) is: $\lambda_{|X} = \sqrt{\frac{\delta^2 P(1-P)Jn}{(1-R_1^2)(1-\rho)}}$ (10)

The statistical power is: $1-\beta = 1 - P[T'(J(n-1)-2, \lambda_{|X}) < t_0] + P[T'(J(n-1)-2, \lambda_{|X}) \le -t_0]$.

The minimum detectable effect size (MDES) regarding the standardized coefficient is:

$$MDES(|\delta|) = M_v \sqrt{\frac{(1-R_1^2)(1-\rho)}{P(1-P)Jn}} \tag{11}$$

where, $M_v = t_\alpha + t_{1-\beta}$ for one-tailed tests with v degrees of freedom ($v = J(n-1)-2$), and $M_v = t_{\alpha/2} + t_{1-\beta}$ for two-tailed tests. The 100*(1-α)% confidence interval for $MDES(|\delta|)$ is given by: $(M_v \pm t_{\alpha/2})\sqrt{\frac{(1-R_1^2)(1-\rho)}{P(1-P)Jn}}$ (12)

Extension to Binary Moderator

When the level-1 moderator, X_{ij}, is a binary variable with a proportion of *Q* in one subgroup and (1-*Q*) in another subgroup, $X_{ij} \sim Bernoulli(Q): VAR(X_{ij}) = \sigma_X^2 = Q(1-Q)$ (13)

For the *random slope model*, we insert Expression 13 into Expression 4.

The noncentrality parameter (unstandardized) is:

$$\lambda_{|X} = \sqrt{\frac{\hat{\gamma}_{11}^2 P(1-P)J}{(1-R_{2T}^2)\omega\tau_{00}^2 + (1-R_1^2)\sigma^2/(nQ(1-Q))}} \tag{14}$$

By standardization, let $\tau_{00}^2 + \sigma^2 = 1$, or let $\delta = \frac{\gamma_{11}}{\sqrt{\tau_{00}^2 + \sigma^2}}$, the noncentrality parameter (standardized) is: $\lambda_{|X} = \sqrt{\frac{\delta^2 P(1-P)J}{(1-R_{2T}^2)\rho\omega + (1-R_1^2)(1-\rho)/(nQ(1-Q))}}$ (15)

The statistical power is: $1-\beta = 1 - P[T^{'}(J-2, \lambda_{|X}) < t_0] + P[T^{'}(J-2, \lambda_{|X}) \le -t_0]$.

The minimum detectable effect size (MDES) regarding Cohen's *d* is:

$$MDES(|\delta|) = M_v \sqrt{\frac{(1-R_{2T}^2)\rho\omega + (1-R_1^2)(1-\rho)/(nQ(1-Q))}{P(1-P)J}} \quad (16)$$

where, $M_v = t_\alpha + t_{1-\beta}$ for one-tailed tests with v degrees of freedom ($v = J-2$), and $M_v = t_{\alpha/2} + t_{1-\beta}$ for two-tailed tests. The 100*(1-*α*)% confidence interval for $MDES(|\delta|)$ is given by: $(M_v \pm t_{\alpha/2})\sqrt{\frac{(1-R_{2T}^2)\rho\omega + (1-R_1^2)(1-\rho)/(nQ(1-Q))}{P(1-P)J}}$ (17)

For the *fixed slope model*, we insert Expression 13 into Expression 9.

The noncentrality parameter (unstandardized) is: $\lambda_{|X} = \sqrt{\frac{\hat{\gamma}_{11}^2 P(1-P)Q(1-Q)Jn}{(1-R_1^2)\sigma^2}}$ (18)

By standardization, let $\tau^2 + \sigma^2 = 1$, or let $\delta = \frac{\gamma_{11}}{\sqrt{\tau_{00}^2 + \sigma^2}}$, the noncentrality parameter (standardized) is: $\lambda_{|X} = \sqrt{\frac{\delta^2 P(1-P)Q(1-Q)Jn}{(1-R_1^2)(1-\rho)}}$ (19)

The statistical power is: $1-\beta = 1 - P[T^{'}(J(n-1)-2, \lambda_{|X}) < t_0] + P[T^{'}(J(n-1)-2, \lambda_{|X}) \le -t_0]$.

This result is consistent with the results about the power formula for the binary level-1 moderator in Spybrook, Kelcey, and Dong (2015).

The minimum detectable effect size (MDES) regarding Cohen's *d* is:

$$MDES(|\delta|) = M_v \sqrt{\frac{(1-R_1^2)(1-\rho)}{P(1-P)Q(1-Q)Jn}} \quad (20)$$

where, $M_v = t_\alpha + t_{1-\beta}$ for one-tailed tests with v degrees of freedom ($v = J(n-1)-2$), and $M_v = t_{\alpha/2} + t_{1-\beta}$ for two-tailed tests. The 100*(1-*α*)% confidence interval for $MDES(|\delta|)$ is given by: $(M_v \pm t_{\alpha/2})\sqrt{\frac{(1-R_1^2)(1-\rho)}{P(1-P)Q(1-Q)Jn}}$ (21)

Results and Conclusions:

The standard error formulas in Expression 3 indicates that the standard error of the moderation effect estimate is not associated with the residual variance ($\tau_{00|T}^2$) for the intercepts, but is associated with the residual variance at level 1 ($\sigma_{|X}^2$). This suggests that adding more covariates in the intercept model would not reduce the standard error or improve power to detect the moderation effect, which is different from the main effect analysis, however, adding more covariates at level-1 that can further explain level-1 variance would reduce the standard error and increase power.

This abstract only shows the partial results in Table 1, we will present more results and demonstrate their application in *PowerUp!* in the presentation.

Appendices

Appendix A. References

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In Howard S. Bloom (editor), *Learning more from social experiments: Evolving analytic approaches*, 115-172, New York: Russell Sage Foundation.

Dong, N. (2014, March). *Power Analysis to Detect the Effects of a Continuous Moderator in 2-Level Simple Cluster Randomized Experiments.* Paper presented at the Spring 2014 Research Conference of the Society for Research on Educational Effectiveness (SREE), Washington, DC.

Dong, N., & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness, 6(1), 24-67.* doi: 10.1080/19345747.2012.673143. PowerUp! is available at: http://web.missouri.edu/~dongn/

Mathieu, J .E., Aguinis, H., Culpepper, S. A., and Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 97* (5), 951-966.

Spybrook, J. (2014). Detecting intervention effects across context: An Examination of the Power of Cluster Randomized Trials. *Journal of Experimental Education*, *82(3), 334-357.*

Spybrook, J., Kelcey, B., & Dong, N. (2015). *Power for detecting treatment by moderator effects in two and three-level cluster randomized trials*. Manuscript submitted for publication.

Table 1: Designs and Analysis Models of Cluster Randomized Trials Covered in This Study

1	2	3	4	5	6
Levels of Clustering =Treatment Assignment Level	Model Number	Level of Moderator	Slope Effect of Lower-Level Moderator	Distribution of Moderators	
2	CRT2-1f	1	Fixed	Binary	Continuous
	CRT2-1r	1	Random	Binary	Continuous
	CRT2-2	2	NA	Binary	Continuous
3	CRT3-1f	1	Fixed	Binary	Continuous
	CRT3-1r	1	Random	Binary	Continuous
	CRT3-2f	2	Fixed	Binary	Continuous
	CRT3-2c	2	Random	Binary	Continuous
	CRT3-3	3	NA	Binary	Continuous